

Predicting Neighbourhood House Price Based on Variety of Economic Indicators

Suran Wu

December 20, 2020

Code and data can be accessed via github at <https://github.com/SuranWu/STA304-Final> , licensed under MIT.

Abstract

Being able to predict house price of neighborhood provides variety of benefits for both the government and individuals. This analysis uses Toronto neighborhood data, trying to build a linear relationship between house price and other economic indicators using multiple linear regression paired with propensity score matching. In the end, the fact that economic indicators Debt risk score and number of social assistance recipients play a important role in predicting house price is found. Other related thoughts, discussion and conclusion are made in this analysis.

Keywords

Statistical Science, Economy, Multiple Linear Regression, Propensity Score Matching, Observational Study, Toronto Neighborhood House Price

Introduction

Residential property has been one of the most important topics for human society regardless of nations or eras. This is exceptionally true for Toronto which is a plural modern city with a large population. Being able to monitor, collect and predict house sale prices are priorities when it comes to discussion about residential property. Statistical ability becomes crucial when we need to collect, clean, and manipulate data from related areas and ultimately allowing us to derive insights from those data. In fact, such an ability is so important that Statistics Canada created a program called The New Housing Price Index (NHPI) which specializes in analyzing the relationship of the price of newly listed and sold houses and other factors. In addition, being able to predict house price allow both government and individual to benefit from it. The government could estimate a variety of things and plan beforehand, property tax income and budget needed for infrastructure for that area for example. For individuals, it may help make decisions on property investment which may promote economic growth. In this analysis, we would try to predict house sale price in different neighborhoods of Toronto based on some economic data available.

The multiple linear regression (MLR) model paired with propensity score matching (PSM) would be used in this analysis. Multiple linear regression can model a linear relationship between the response variable (sale price) and other predictors. So the model, analysis, and other diagnostic sections would be performed, empowering us to predict new house prices with the knowledge gained from some other economic data (Sheather, 2008). This analysis is based on observational study, propensity score matching is a statistical matching technique, allowing us to reduce bias due to confounding variable. The method attempts to estimate the effect of treatment (in our case, whether a neighborhood with high debit risk score) by accounting for covariates of getting treatment (Rosenbaum and Rubin, 1983). An MLR model is appropriate since there is a linear relationship between the response and explanatory variables. More on the appropriateness and validness of the model would be discussed in the later parts of this analysis.

For explanatory variables, one may find them are all good indicators of neighborhood wealth and therefore house prices can be predicted. For example, Citizens who can afford to pay for daycare for their children may tend to be wealthier and live in more expensive homes. A positive correlation may be speculated by intuition. So it would be an interesting model to use those variables to predict house prices for a specific neighborhood. More explanation of variables would be provided in the Data section.

Throughout the whole analysis, for the Data section, it will provide a detailed explanation, method, and measurement on each variable. After performing propensity score matching to clean the data, the MLR model will be fitted in the model part. The estimated value from the model would be presented in the result part. In the discussion section, we will talk about the summary, analysis, diagnostic discussion. Other potential inferences along with conclusions would be drawn as well. Finally, weaknesses and next steps will be analyzed and suggested.

Data

The table 1 which is baseline characteristics table is presented as following. It gives a glance of overall statistical information.

Table 1 : baseline characteristics table

	Neighborhood without high debt risk score (N=107)	Neighborhood with high debt risk score (N=33)	Overall (N=140)
Businesses			
Mean (SD)	587 (694)	372 (366)	536 (637)
Median [Min, Max]	368 [47.0, 4320]	313 [58.0, 2090]	346 [47.0, 4320]
Child.Care.Spaces			
Mean (SD)	117 (79.9)	150 (83.3)	125 (81.6)
Median [Min, Max]	105 [0, 441]	142 [0, 367]	110 [0, 441]
Debt.Risk.Score			
Mean (SD)	728 (23.0)	775 (8.71)	739 (28.6)
Median [Min, Max]	730 [661, 760]	775 [761, 793]	741 [661, 793]
Home.Prices			
Mean (SD)	469000 (160000)	805000 (370000)	548000 (268000)
Median [Min, Max]	458000 [204000, 993000]	752000 [375000, 1850000]	491000 [204000, 1850000]
Local.Employment			
Mean (SD)	10400 (21600)	6200 (5790)	9410 (19100)
Median [Min, Max]	4070 [438, 186000]	4030 [924, 25600]	4050 [438, 186000]
Social.Assistance.Recipients			
Mean (SD)	2250 (1500)	618 (506)	1860 (1500)
Median [Min, Max]	1790 [189, 7260]	503 [37.0, 1930]	1460 [37.0, 7260]

Key components and detailed descriptions of each variable are presented as following:

Home price : House price (Average price for residential real estate sales) is collected during the period 2011-2012 and it is collected by Realosophy.com. It is collected by third party and used by Canadian government. The target population would be all houses listed and sold throughout Toronto area in that year. The frame would be related government and company documents that record all house prices listed and sold of Toronto in that year. Sample would be the same as frame as it is considered as a census. Strength of this variable would be it provides a trusted, accurate and well-rounded data since it is objective data collected and used by the government. Weakness would be it ignores inequality issue. For example, for a same neighborhood, house prices could be huge in difference because of floor plan, available areas and many other factor.

Businesses : It represents how many business licensing granted for that neighborhood as 2011. It is recorded by City of Toronto, Municipal Licensing & Standards. It is objective data collected by the government instead of a statistical survey. It should be considered as a census, target population would be all business license issued in that neighborhood. Frame would be based on related documents of City of Toronto, Municipal Licensing & Standards. Sample would be same as frame.

Child Care Space : It represents licensed Child Care Spaces for a specific neighborhood. Methodology of how data collected, related discussion of target population, frame and sample is the same as Businesses variable. Strength of the data would be it is a good economic indicator of neighborhood wealth, as discussed in Introduction section. Weakness of this data would be approximately 1/3 of families choose child care outside their neighborhood (City of Toronto, 2011). Some neighborhoods tend to have more spaces as commuters bring their children to centers near their work or along their commuting route.

Local Employment : It represents total employment (workers with age 15+) in certain neighborhood. It is collected by Research and Information unit of city planning, City of Toronto. Methodology of how data collected, related discussion of target population, frame and sample is the same as Businesses variable. Weakness would be it is collected based on individual worker who report their working activity by tax form by themselves or their employees. So it tends to underestimate the total employment because of many illegal working.

Social Assistance Recipients : It represents count (by members, not cases) of recipients of aid qualifying for Ontario worker in that neighborhood. The Source would be from Toronto Employment & Social Services. Methodology of how data collected, related discussion of target population, frame and sample is the same as Businesses variable. Weakness would be some qualified workers may not know certain aid program and this tends to underestimate workers qualified for that neighborhood.

Debt Risk Score : It is proprietary index value provided by TransUnion Canada that indicates the likelihood of missing three consecutive loan payments. It is measured under a systematic scoring mechanism. Debt risk score is collected by TransUnion Canada in 2013 with the census method. TransUnion data is provided by postal code and covers approximately 92% of all Canadians with credit files. The remaining 8% is dropped due to privacy reason. Weakness would be postal codes with fewer than 15 credit files are suppressed (TransUnion 2013) so it tends to have selection biased. In the end, high score would mean those score bigger than 760 (TransUnion 2013). And this is the standard for 'treatment' in this analysis for defining neighborhood with high debt risk score or not

To visualize the data, scatter plots matrix of each variable are presented as following.

Figure 1 : Scatterplot Matrix Between Response and Explanatory Variables

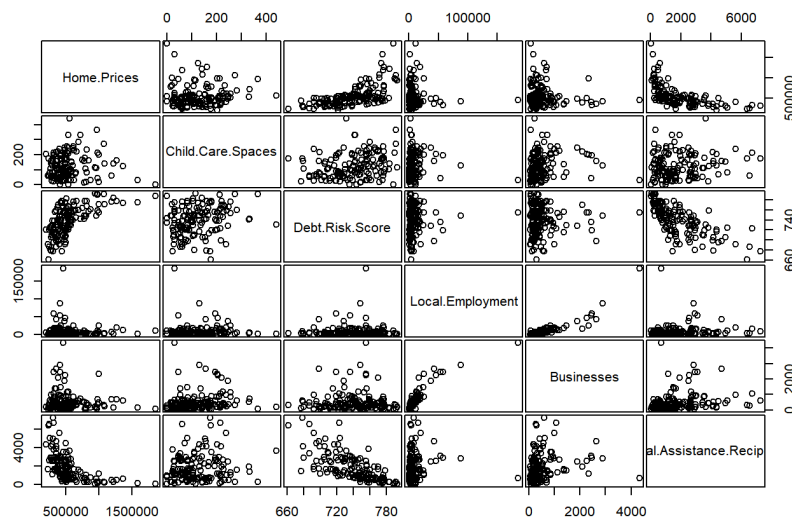


Table 2 : Pairwise Correlation Between Explanatory variables

	Child Care Space	Debt Risk Score	Local Employment	Businesses	Social Assisatance Recipient
Child Care Space	1.0000	0.2435	0.1238	0.1402	0.0074
Debt Risk Score	0.2435	1.0000	-0.6971	0.0053	0.0451
Local Employment	0.1238	-0.6971	1.0000	0.1812	0.0754
Businesses	0.1402	0.0053	0.1812	1.0000	0.8803
Social Assisatance Recipient	0.0074	0.0451	0.0754	0.8803	1.0000

In addition, due to concern of multicollinearity for using MLR model, pairwise correlation between each of explanatory variables is shown as the Table 3. Multicollinearity is a situation in which two or more explanatory variables in a multiple regression model are highly linearly related. It is a huge problem when we use MLR model because it undermines the statistical significance of an independent variable. In general, an absolute correlation coefficient of >0.7 among two or more predictors indicates the presence of multicollinearity (RekhaMolala, 2019). As table suggests, correlation between Businesses variable and Local Employment variables is estimated as 0.8803 which indicates problem of multicollinearity. So, we would need to remove one of them from our explanatory variable. In the end, Local Employment is dropped due to weakness of this variable. Local Employment tends to underestimate from the true value while Businesses data is obtained by trusted source and it is accurate by the nature of measurement for this variable as discussed above.

Model

A propensity score matching (PSM) is performed before the multiple linear regression (MLR) is fitted. In this analysis, neighborhoods with high debt risk score are considered as 'treatment group' as discussed in Data section. Therefore, a logistic regression model is performed to calculate fitted value for likelihood of whether a neighborhood has high score. Then, a nearest neighborhood matching is performed based on fitted value obtained from the logistic model. A balance between treatment and comparison group on observable traits is observed. This analysis follows the standard procedures for propensity score matching and cleaned data set is obtained.

The multiple linear regression is fitted based on new data set. From the scatter plots matrix presented in Data section, we may observe strong linear relationship between our response variable home price and other explanatory variables. So the MLR could be used as we want to predict the value of a variable (home price in our case) based on the value of two or more other variables. In multiple linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. In addition, Multiple linear regression is very useful because it can examine which variables are significant predictors of the outcome variable. We may drop explanatory variable which play less significant role in predicting our response variable. Since our goal is prediction, the MLR can be used to fit a predictive model based on known data. If we have additional information about other explanatory variables, we may use this model to make prediction of the response.

Since all of our explanatory variables are good economic indicators for a certain neighborhood, they should have the ability to predict house price in that neighborhood. The explanatory variables are either objective data collected by the government or a score obtained under a systematic scoring mechanism, no other manipulations are conducted besides procedure needed for propensity score matching. So data features are explained in Data section and this analysis directly uses the data set as is.

All of this analysis, including but not limited to propensity score matching, MLR model fitting, graph plotting, are performed by R Studio.

Expected MLR model equation would be as follows:

$$\hat{y} = b_0 + b_1X_B + b_2X_C + b_3X_D + b_4X_A$$

\hat{y} : Expected home price

b_0 : Estimated intercept

b_1 : Estimated slope parameter for Businesses variable

X_B : Input value of variable Businesses, ranged from 47 to 4320.

b_2 : Estimated slope parameter for Child Care Spaces variable

X_C : Input value of variable Child Care Spaces, ranged from 0 to 441.

b_3 : Estimated slope parameter for Debt Risk Score variable

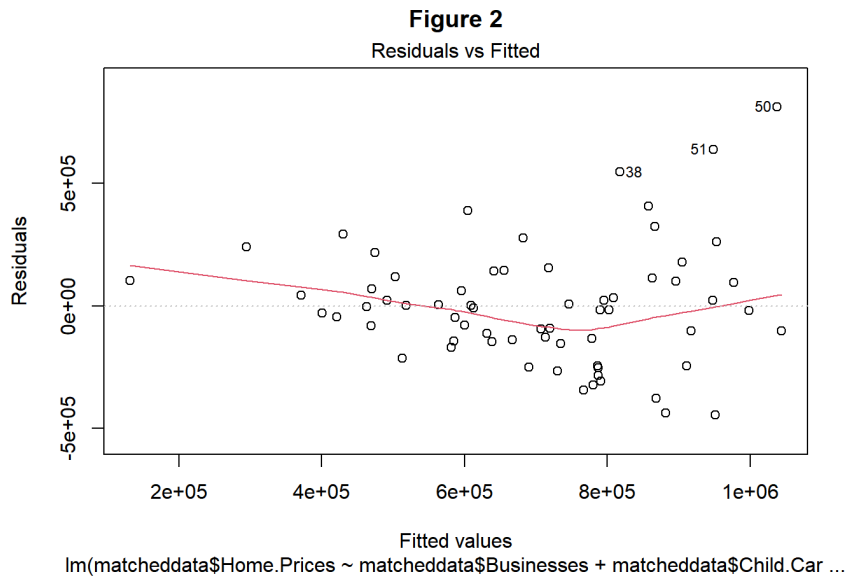
X_D : Input value of variable Debt Risk Score, ranged from 661 to 793.

b_4 : Estimated slope parameter for Social Assistance Recipients variable

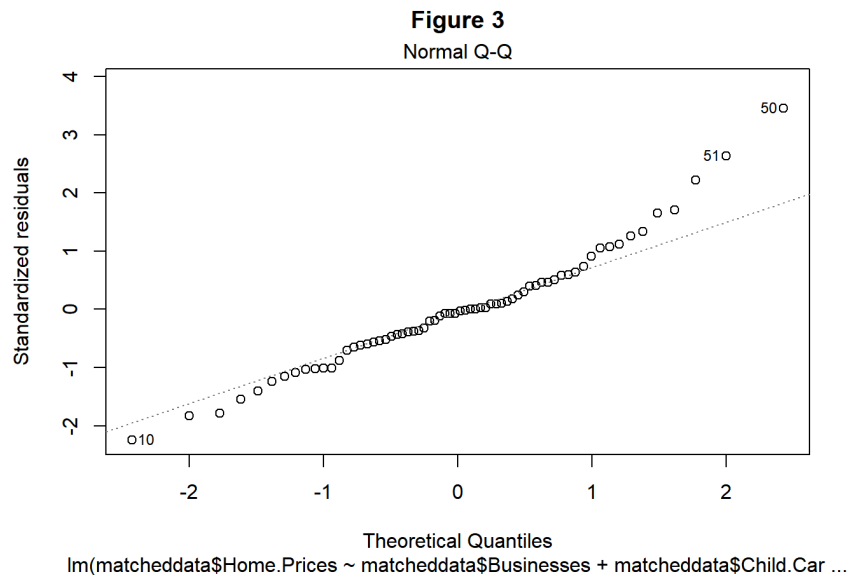
X_A : Input value of variable Social Assistance Recipients, ranged from 37 to 7260.

Model Check: Here are 4 standard diagnostic plots created for the sake of validating the model.

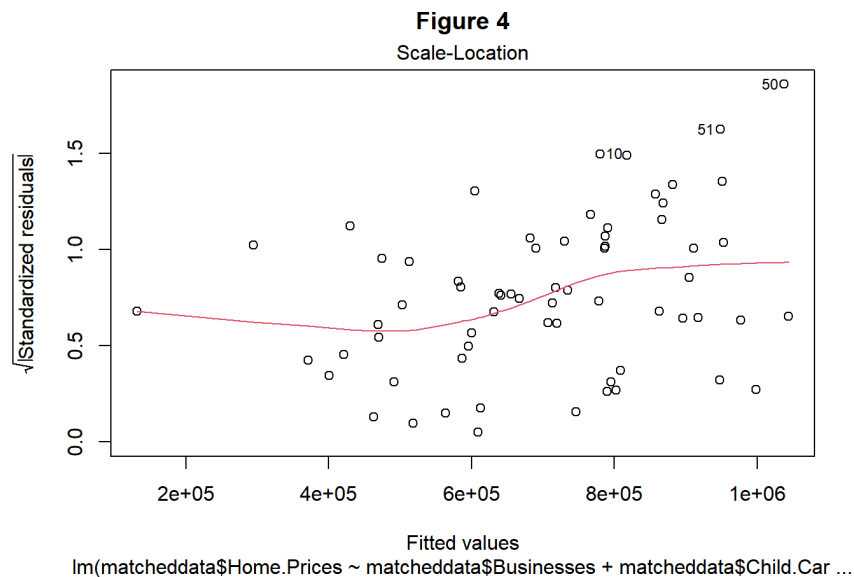
Keep in mind that our total observation is 140 (140 neighborhood in Toronto suggested by the government). After using propensity score matching, the available observation is reduced to 66. This small sample size may cause difficulties when it comes to model check.



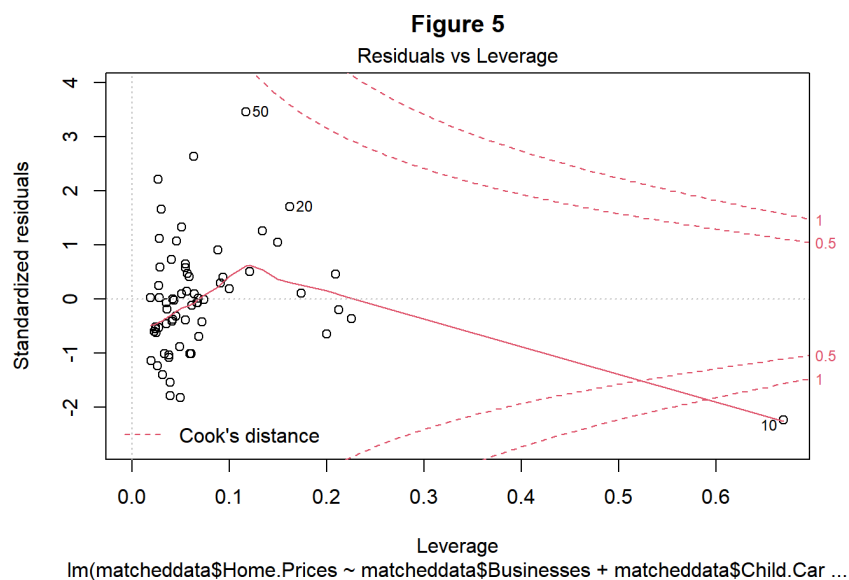
For the first plot, it is residual vs fitted value plot. It checks linear relationship between explanatory variables and response variable. We can see that residuals are almost randomly and about equally spread out around horizontal line without distinct pattern as the red line is relatively flat. Considering the small sample size we have, we may tolerate the little pattern (the red line decreases with a small slope and increases with a small slope after). What's more, the residuals bounce around horizontal x-axis which means their mean are about 0. Thus, we may conclude we do have a linear relationship between our explanatory variables and response variable. Our model makes sense at very least.



For the second plot, it is normal Q-Q plot of standardize residuals. It checks normality assumption of residuals. From the graph, we can see most of the standardize residuals are in the line. However, we may also observe it is a bit of heavy tails. That is, in extreme cases, the small value is smaller than expected and large value is larger than expected. So except for some extreme cases, like case 10, 51 and 50, we may conclude normality assumption is met.



For the third plot, it is standardized residuals vs fitted value. It checks the equal variance assumption (homoscedasticity). From the plot, we see the red line is about horizontally and standardize residuals are almost randomly placed around the horizontal red line. Again, considering small sample size, we may tolerate the little pattern of the red line. No distinct and noticeable pattern found in this plot, from that, we may conclude equal variance assumption is met.



For the last graph, it is plot of residuals vs leverage. It helps us to find any influential cases. This is important as high leverage points are not necessarily being negatively influential to the whole model. From our plot, we can spot one case that has high leverage and it is out of range from the cook's distance. So the case 10 may be a influential point of the model and we may want to remove it. Other than case 10, we find no point that is really beyond cook's distance, so no other influential points found.

Results

An additive multiple linear regression is fitted from the Model section and following table presents the estimated coefficients, standard error, t value and their p-value for t-test.

Table 3 : Estimated Coefficients Summary

	Intercept	Businesses	Child Care Spaces	Debt Risk Score	Social Assistance Recipients
Estimated Coefficient	-2801000	1.741	-9.871	4883	-26.25
Std. Error	1613000	5.105	424.8	2108	8.68
T value	-1.736	0.341	-0.232	2.316	-3.025
P-value (For t-test)	0.08757	0.73422	0.81703	p<0.05	p<0.05

So the Fitted equation of this MLR model would be :

$$\hat{y} = -2801000 + 1.741X_B - 9.871X_C + 4883X_D - 26.25X_A$$

Where \hat{y} : Expected home price

X_B : Input value of variable Businesses, ranged from 47 to 4320.

X_C : Input value of variable Child Care Spaces, ranged from 0 to 441.

X_D : Input value of variable Debt Risk Score, ranged from 661 to 793.

X_A : Input value of variable Social Assistance Recipients, ranged from 37 to 7260.

In addition, we may find that the model suggests only variable Debt Risk Score and Social Assistance Recipients are significant since their p value is less than 0.05. This is a strong evidence that we reject the null hypothesis that the slopes of those 2 variable are 0 and it suggests strong linear relationship may be found. However, for variable Businesses and Child Care space, the p value is greater than 0.05, so we fail to reject the null hypothesis that the slopes of those 2 variable are 0. It suggests our data do not support a strong linear relationship between home price with those 2 variables.

Alternatively, if we rebuild our model by backward elimination with AIC, the estimated coefficients summary are presented as following.

Table 4 : Estimated Coefficients Summary for AIC method

	Intercept	Debt Risk Score	Social Assistance Recipients
Estimated Coefficient	-2686685.5	4728.3	-265.8
Std. Error	1399943.4	1804.5	72.8
T value	-1.919	2.62	-3.651
P-value (For t-test)	p<0.05	p<0.05	p<0.05

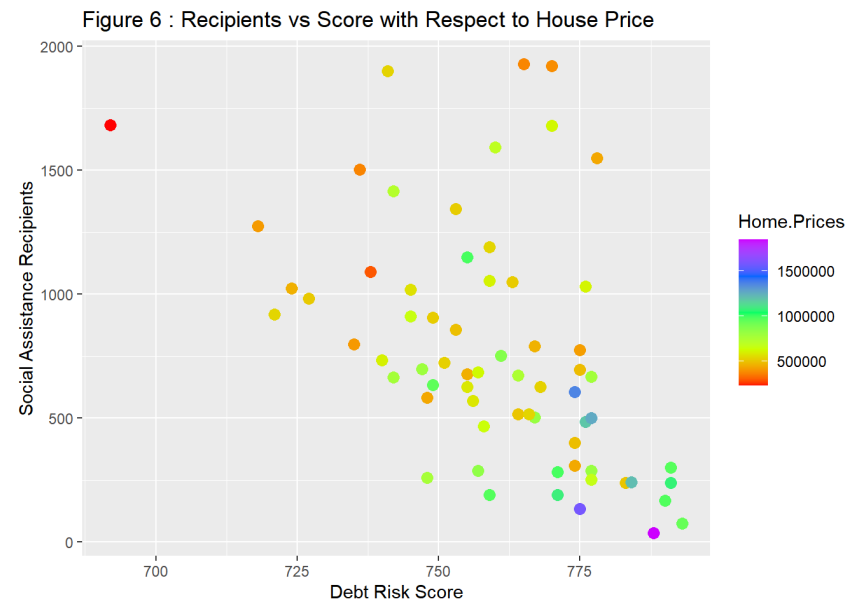
And by results obtained by back elimination with AIC, our equation becomes

$$\hat{y} = -2686685.5 + 4728.3X_D - 265.8X_A$$

X_D : Input value of variable Debt Risk Score, ranged from 661 to 793.

X_A : Input value of variable Social Assistance Recipients, ranged from 37 to 7260.

To visualize what our result is, the following plot is created.



This plot describe how Debt Risk Score and Social Assistance Recipients together would affect home price in neighborhoods of Toronto. From the summary table, we see that house price has positive relation with debt risk score and negative relation with number of social assistance recipients. And this plot would help us double check with the results obtained by model.

As we can see, house prices are represented by color of the point. The more expensive the price is, the point will be more like blue-purple-ish color and the less expensive, the color of point will be more of yellow to red. From the plot, we can see points in top left, with high y value and low x value (therefore, larger number of social assistance recipients and lower debt risk score) will tend to be yellow to red (so less expensive house price in that area). In contrast, points in bottom right, with low y value and high x value (therefore, smaller number of social assistance recipients and higher debt risk score) will tend to be more of green to blue to purple-ish color. This transition of color of the points from top left to bottom right is consistent with what we have in summary table of estimates coefficients.

Discussion

1 Summary of this analysis:

The goal to this analysis is to build a model to predict Toronto house price on different neighborhood based on some economic indicators, namely, Debt risk score, Businesses, Social assistance recipients and Child care space. We achieve this goal by following steps. Firstly, we obtain the data from open data distributed by the City of Toronto. The variables are either objective data collected by the government or a score obtained under a systematic scoring mechanism. To maintain objectivity and avoid publication bias, no other special manipulations are done to clean the data expect for steps that required by statistical study. Then, a propensity score matching is performed to matched the neighborhoods based on likelihood of whether a area is likely to be high score area. This step helps us to creates a balanced dataset, allowing a simple and direct comparison of baseline covariates between area with or without high score(>760 as suggested). Finally, the multiple linear regression model is fitted based on this balanced dataset and it allows us to obtain our predictive equation. In addition, we also examine validness of our model and conclude it meets most of the assumptions required.

2 Interpret Result:

Our predictive equation suggested by the MLR model would be

$$\hat{y} = -2801000 + 1.741X_B - 9.871X_C + 4883X_D - 26.25X_A$$

The interpretation would be, set debt risk score as a example, For debt risk score variable, if holding other predictors fixed, when the score increased by one in that neighborhood, we would expect sale price to increase by 4883 in average. Interpretation of other variables follows above. If we know a neighborhood with 300 business licenses, 100 Child care space, 1000 social assistance recipients and 760 debt risk score, our predicted house price of that area would be $-2801000 + 1.741 * 300 - 9.871 * 100 + 4883 * 760 - 26.25 * 1000 = 883370.6$.

However, by observing p value of variable Businesses and Child Care Space, we may conclude those 2 variables are not significant on predicting house price based on our balance model, so we may want to drop such variables. The commonly-used method to find a parsimonious model is stepwise regression with AIC. In backward elimination, it starts with all the potential predictors in the model, then removes the predictor with the largest p-value each time to give a smaller AIC. If we use such a method, trying to drop non-significant explanatory variables, the final model would become

$$\hat{y} = -268668.5 + 4728.3X_D - 265.8X_A$$

3 Significance:

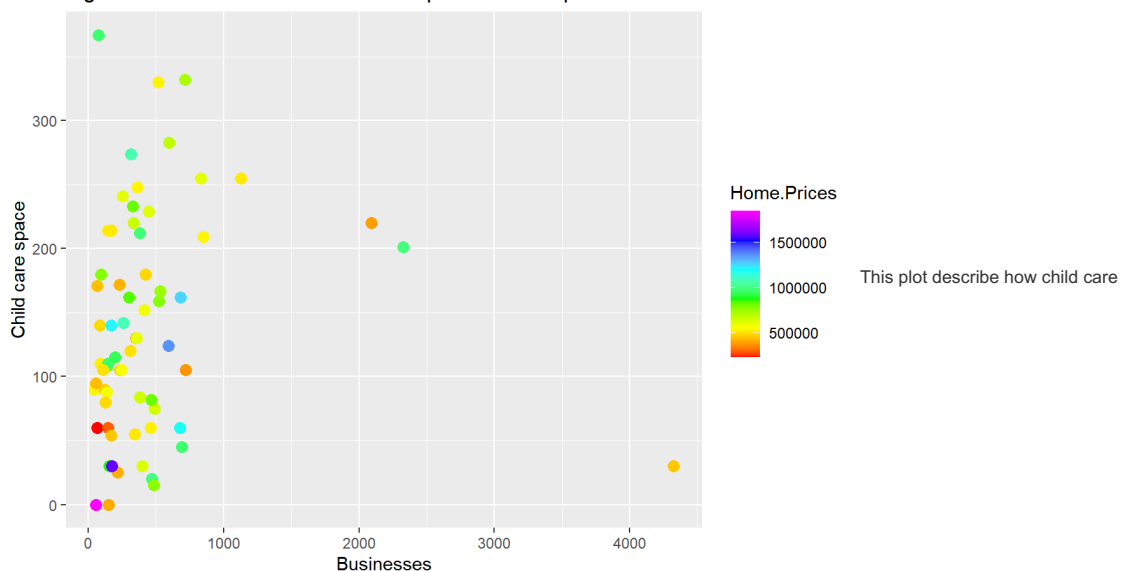
This analysis which use MLR model paired with PSM method shows that price of house in neighborhood of Toronto likely depends on average debt risk score of residents how many social assistance recipients in that neighborhood. If a new property is build in certain area, the government will be able to estimate a variety of things and plan beforehand, property tax income and budget needed for infrastructure for that area for example. Both government and citizens would benefit from it. In addition, this model suggests potential relationship exists between property price and other economic indicators of certain area in Toronto, governments with different cities or even countries could use this model as a reference to explore what local economic indicators have effect on house sale price. This may help social planner for macro-control on house market and other area.

4 More on Visualizing Results

In the Result section, we verify the model by plotting Figure 6 : Recipients vs Score with Respect to House Price. And we success in observing the transition of color which is consistent with the result from model equation. Please review Section part for the figure and related explanation.

Now, we would visually check why other 2 variables which are Businesses and Child Care Space is not significant in predicting house price:

Figure 7 : Businesses vs Child care space with Respect to House Price



space and businesses together would affect home price in neighborhoods of Toronto. Those 2 explanatory variables are considered as not significant factor in predicting house price. This time, we fail to see any distinct pattern of transition of color which indicates house price as same

layout with figure 6. So no distinct transition of color would suggest those 2 variables are not that useful when we try to study their relation with house price. So the AIC model we get which drops these 2 variables makes sense.

5 Weakness

There are some weakness exists through this analysis. Firstly, there are weakness from data itself. For example, house price and debt risk score are collected based on different year which is questionable. For house price data, it fails to consider equality problem, house prices could be huge in difference. This is particularly true for areas with condominium, with different floor plan and area of the condo, price of some units could even be double than the others in same neighborhood. Also, the data based on neighborhood is obtained by the mathematical aggregation of smaller sub-areas (in this case Census Tracts) that when combined, define the entire neighborhood. Since smaller areas are more likely to have their values rounded or suppressed, the overall total may be under-counted.

When it comes to method, propensity score matching is used when we try to clean a balanced data. The new data does not contain all neighborhoods, and, therefore some information about the original dataset structure is lost. This is more of a problem for this analysis, there are originally 140 neighborhoods observation. After matching, it is reduced to 66 observations which is considered as small sample size. This would cause difficulties when we fit the model and extreme cases are more likely to happen.

6 Nest Steps

To overcome small sample size problem, I would suggest obtain data not by neighborhood but by divisions of street. This would allows us to have much more sample cases. What's more, we could leave neighborhood to match with divisions of street, this allows us to do cross validation which will verify accuracy of the model. Finally, this may increase the model accuracy because of the nature of house price market overall.

Secondly, we may want to compare house price of the 'same kind' so the inequality issue discussed in weakness could be eliminated. More specifically, we want to only compare townhouse to townhouse or Condo with one bedroom to another, inside one specific neighborhood. Finally, we may want to compare the results with different method of the model. For example, what would the model be if we use BIC criterion? Maybe build the model by penalized regression? what about fit the model without propensity score matching? We could perform those and compare statistical results (like adjusted R square) from there.

Reference

1. Social Development, Finance & Administration. 'About Wellbeing Toronto - Economics' Toronto Open Data. <https://open.toronto.ca/dataset/wellbeing-toronto-economics/> 2014. 12-31
2. Sheather, Simon J, and George Casella, and Stephen Fienberg, and Ingram Olkin. "Simple linear regression." A Modern Approach to Regression with R. Springer, 2008. 10
3. Molala, Rekha. 'MLmuse: Correlation and Collinearity — How they can make or break a model', Clairvoyant. <https://blog.clairvoyantsoft.com/correlation-and-collinearity-how-they-can-make-or-break-a-model-9135f6e6936a>, 2019. 7-15.
4. Nuttall, Gregory A, and Houle, Timothy T. 'Liars, Damn Liars, and Propensity Scores.' Anesthesiology. Vol. 180 3-4. <https://doi.org/10.1097/01.anes.0000296718.35703.20>, 2018. 1.
5. Littnerova, Simona, and Jarkovsky, Jiri, and Pavlik, Tomas. 'Why to use propensity score in observational studies? Case study based on data from the Czech clinical database AHEAD 2006–09' Cor et Vasa, vol. 55 4. <https://doi.org/10.1016/j.crvasa.2013.04.001>, 2013. 8.
6. Alexander, Rohan. 'Difference in differences', Telling Stories With Data. https://www.tellingstorieswithdata.com/06-03-matching_and_differences.html, 2020. 11-5.
7. Rich, Benjamin. Table1 : Package 'table1', R package version 1.2.1. <https://github.com/benjaminrich/table1>, 2020. 11-19.
8. Rosenbaum, Pual R, and Rubin, Donald B. The central role of the propensity score in observational studies for causal effects, Biometrika. Vol. 70 1. https://watermark.silverchair.com/70-1-41.pdf?token=AQECAHi208BE49Ooan9kKhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAaQJwggKhBgkqhkiG9w0BBwagggKSMiICjglBADCCAocGCSqGSib3DQEHATAeBgIghkgBZQMEAS4wEQQMHb-URNcMUUpNvbzXvAgEQgllCWEKjX51PYZ14BOjEPzRMO0cKfrMwqti
i9tITIFsaCCbkZBVGmuLpw9KO9hu_0aydi_FysFsFBfn7K6JPI5FZeyHUqBGmSqK-Fa_m5AOQdQ6mliHtS8RbecxQbzJE96B0Fhpy4ogMsorlnjPE9xgzSjY-BSVhrutnH8TY799Zh29XIels5xdNmQjb7M_I6qoovCAnbR0agnHn23SGKKGI_SXf9A6i4Pvy1bsdV4-z-IM3uXI3_8_LQkqfXPBTnievMI9chhS9dW1sfJ7Te0u0B799oKgmZ7K6vxd2zsS9nF8G8eaM7pSRMqLdE1PqZ1CCcj5BuGVaPO4hhCroCirAeeEs523ytLe6S8GkajFBm6CNc0eWbywuAn9ZMaGGbC7GDN7GhFS2QVKEz28zagkcLbWfolPpTGwYAN-V7AEhUMmHL1aIOuQjAlkfeeh0knmBODjBXrivmeRYIdhDkKqpE5rF3iayW2c8C6KmFFo9QS9XYk8BKD9jnsQHOluwNmTdxdhnh0KuX240jHRhf4TDhgUqgLo4oDwykdIWc4C0EzJLtgymWffSu57poc_Ovqi_ExWqIScUVLd5ukauufGHv3S8p4h1O_GJxJwdypyt_X5mN-NeHETf2N8vyDurH2id4PUIKUhz1j83hX7nfcNsGNctmCgNcOKwNWaM8tZw89WxlMBccAhHzLVaqXEGsGSnTKI3MdQOVWNLqr4igB2OYU8TmmEyH7puRQB1j2Q7TujtZzAjqLMN4Zzpc33pVo38MHKO8BF4xjiA. 1983.

Loading [MathJax]/jax/output/HTML-CSS/jax.js New Housing Price Index (NHPI). <https://www.statcan.gc.ca/eng/survey/business/2310>. 2020.